Article:

Madigan, R orcid.org/0000-0002-9737-8012 and Romano, R orcid.org/0000-0002-2132-4077 (2020) Does the use of a head mounted display increase the success of risk awareness and perception training (RAPT) for drivers? Applied Ergonomics: human factors in technology and society, 85. 103076. ISSN 0003-6870

https://doi.org/10.1016/j.apergo.2020.103076

Does the use of a head mounted display increase the success of risk awareness and perception training (RAPT) for drivers?

Ruth Madigan[a] and Richard Romano[a]

[a] *Institute for Transport Studies, University of Leeds, LS2 9JT, United Kingdom;*

Abstract

The PC-based driver training programme, Risk Awareness and Perception Training (RAPT) has been successful in improving young drivers' hazard anticipation and mitigation responses in both simulator and on-road studies. The current research aimed to evaluate the success of an adaptation of this training for the UK context, along with investigating the impact of the presentation modality on RAPT effectiveness. Traditionally RAPT has been delivered on a PC monitor, which does not allow the same range of head and eye movements that drivers use when on the road. Thus, it was anticipated that the 360° field-of-view provided by Head Mounted Display (HMD) technology would provide a more ecologically valid experience, facilitating deeper processing and encoding of driving relevant scanning patterns, and an increased capacity to identify potentially hazardous areas of a driving scenario. Using a between-subjects design, three different training modalities were compared - a PC-based version using still images (PC-Stills), a HMD version using still images (HMD-Stills), and a HMD version using videos (HMD-video). All three training groups' performance on the UK Hazard Perception test was compared to that of a control group, who received no training. Results indicated that the adaptation of the training materials for the UK context was successful, with all three training programmes leading to performance improvements in the RAPT tests. Although participants in the HMD-video condition required more attempts to pass the training, this group showed the greatest improvement in hazard perception scores from the pre- to the post-training tests. Results also showed scenario-based differences between the modalities, suggesting that the success of different versions of RAPT may be linked to the type of risky scenario being targeted.

Keywords: Hazard anticipation; Risk awareness; Driver training; Head Mounted Display

1. Introduction

Young, novice drivers have been shown to be one of the most at-risk groups on the road, with both age and inexperience making distinct contributions to high accident rates (McCartt et al., 2003). Poor hazard anticipation, or an inability to scan effectively for potential roadway risks, has been identified as a key skill deficit linked to the increased crash risk, leading to numerous attempts to improve this skill through training (McDonald et al., 2015). The pc-based Risk Awareness and Perception Training (RAPT) developed at the University of Massachusetts (Fisher et al., 2003) has been one of the most promising programmes in this area, with both simulator (Pollatsek et al., 2006) and on-road evaluations (Pradhan et al., 2009) showing evidence of improvements in driver eye-movement patterns and hazard mitigation responses, both in the near and long term (Taylor et al., 2011). Encouragingly, a recent, large-scale study of over 5000 young drivers in California found that RAPT was linked to a 23.7% reduction in the crash rate of young males who had completed the training, when compared to a control group (Thomas et al., 2016). A similar evaluation of RAPT, focusing on novice drivers with less than 3 years' experience, showed a reduction in crash rate of 24.6% (Zhang et al., 2016).

Research into training effectiveness differentiates between near transfer - the application of learning to a similar context to the one in which training takes place, and far transfer - the application of more general learning principles to a dissimilar context (Barnett & Ceci, 2002). As it is impossible to expose learner drivers to all of the potentially hazardous situations they will encounter (Groeger & Banks, 2007), the development of higher order processing skills such as hazard anticipation, are of the upmost importance in facilitating far transfer. Across a variety of domains, error training has been shown to be one of the most effective mechanisms for encouraging far transfer of learning (Keith & Frese, 2008), as the opportunity to make errors and learn from these, facilitates the development of learners' metacognition and problem solving skills, and thus their ability to apply their learning to a wide array of situations. The process of learning during RAPT is based on an error-based feedback mechanism known as the 3M approach, whereby trainees are offered a chance to make a mistake, followed by the offer of an appropriate mediation strategy to fix the mistake, and then an opportunity to master the mediation strategy (Unverricht et al., 2018).

The most recent version of RAPT (RAPT3) consists of nine driving scenarios where latent hazards are hidden from view either due to the geometry of the roadway or the presence of an obscuring vehicle, along with hazards caused by visible elements, either cars or pedestrians that might plausibly enter the drivers pathway (Pradhan et al., 2009). Trainees are provided with top-down schematic representations of driving scenarios, which highlight the potentially risky areas where particular attention should be paid (see Figure 1). The purpose of the top-down views is to encourage novice drivers to reason spatially about a scenario more actively than they would if they saw perspective drawings or actual videos (Pollatsek et al., 2006), thus leading to a greater ability to apply their learning to a variety of situations. Following the presentation of each schematic, trainees are presented with a set of sequentially advancing perspective snapshots of a scenario, using a mouse-click to identify the areas of each image that they would check for latent hazards while driving. Studies testing the effectiveness of RAPT in an on-road setting have found that a higher percentage of trained drivers (64.4%) correctly fixated target zones than untrained drivers (37.4%, Pradhan et al., 2009).

Figure 1: Example of RAPT schematic and image presentation (see www.ecs.umass/hpl)

1.1 Adaptation for UK Context

The success of RAPT in improving young drivers' crash risk in the U.S. shows the promise of this type of training. However, the driving conditions and legal requirements for driving in the U.K. differ considerably from the U.S., with U.K. drivers driving on the opposite side of the road for a start. Therefore, it is important to explore whether this type of training would be equally successful in increasing UK drivers understanding of the types of hazard typically encountered on their roads. Crundall et al. (2012) conducted research in a UK context, separating out different types of latent and visible hazards, and distinguishing between behavioural and environmental prediction hazards.

Behavioural prediction hazards included situations where drivers could predict a hazardous situation by identifying a potential hazard e.g. a pedestrian standing at the side of the road, and assess its attitude, position, and trajectory. Environmental prediction hazards included situations where drivers could identify that the environment through which they were driving may contain a hidden hazard (e.g. a parked truck may be hiding a pedestrian). Their simulator study found that novice drivers were particularly poor at noticing environmental prediction hazards and were likely to miss these altogether. This result shows that the types of latent hazards used in RAPT are also likely to be relevant in the UK context, while also supporting an idea put forth by Sagberg and Bjornskau (2006) that different hazards may be more/less successful in discriminating between experience groups. Thus, the first aim of the current research is to adapt the US-based RAPT materials for use in a UK context, paying particular attention to the specific environmental cues which untrained drivers are weakest at noticing.

1.2 Impact of field of view

Despite the success of RAPT in improving trainees' scanning behaviours and hazard mitigation performance, it has not led to a ceiling effect, showing that there are still improvements to be made. It has been argued that this may be due to a failure to fully cognitively map the knowledge gained from RAPT to a dynamic road environment (Yamani et al., 2017; Agrawal et al., 2018). One potential explanation for this may be that the eye movements taught during the PC-based version of RAPT do not match those required while driving in a dynamic 360° environment on the road, thus requiring far transfer of training on multiple domains (differences in field of vision available, and differences in the specific scenarios encountered), from the learning to the application contexts (Groeger & Banks, 2007). Therefore, one of the aims of this study is to examine the impact of providing a wider field of view on trainees' RAPT performance, through the use of a head mounted display (HMD). It might be expected that the eye movements required in this type of immersive setting would more closely mimic those required during on-road driving, and thus aid the process of cognitive mapping more effectively.

HMDs provide a cheaper alternative to driving simulators, enabling an immersive and engaging experience, and allowing the measurement of head movements while moving through a virtual scene. Agrawal et al. (2018) have shown the success of this type of platform in improving trainees' hazard anticipation skills, when a combined hazard anticipation and hazard mitigation training strategy was used. In their study trainees exposed to a virtual reality version of RAPT (V-RAPT), presented through a HMD, anticipated a greater proportion of latent hazards than traditional RAPT trainees in a subsequent simulator study. However, although the same hazards were used in both training programmes, the delivery of the V-RAPT training materials differed considerably from RAPT, with virtual reality videos being used rather than static photographic images, and training instructions delivered through audio rather than text. Therefore, it is not clear whether performance improvements were a result of the increased head movement facilitated by the training platform, or the changes made to the RAPT training materials. The current study will address this issue, through comparing the effects of displaying 360° static images through the HMD with the effects of using video displays.

1.3 Hazard Perception Testing

As research has shown a link between hazard perception skill, or "the ability to read the road and anticipate forthcoming events" (McKenna et al., 2006, p.2), and crash involvement (Horswill &

McKenna, 2004), many jurisdictions worldwide, including the UK, have incorporated a hazard perception test (HPT) into their licensing procedure. The UK HPT requires participants to watch a series of video clips and respond with a button press when a hazard is detected (Jackson et al., 2009). All of the hazards develop into near miss situations, and research has shown that drivers who have passed the HPT have, on average, a lower accident liability than those who did not take the test (Wells et al., 2008). Therefore, the HPT provides a validated measure of novice driver behavioural hazard anticipation, which will be used in this study to evaluate the effectiveness of the RAPT programmes in improving trainees' capability to understand potentially hazardous road scenarios. It should be noted that because previous research has differentiated between the environmental hazard anticipation trained using RAPT and behavioural prediction hazards measured using HPT (e.g. Crundall et al., 2012), it is not clear if the RAPT training will lead to the far transfer of learning that would be required for improvements to occur in HPT performance.

1.4 Current study

The current study has two main aims. Firstly, it aims to evaluate the success of an adaptation of RAPT3 for the UK context (RAPT-UK), as measured through improvements in performance on RAPT-UK scenarios after training, along with improvements in HPT performance. The second aim is to understand the impact of incorporating a 360° field-of-view on trainees' ability to identify potentially hazardous zones of a scenario.

Three different adaptations of the RAPT-UK will be implemented:

- A PC-based version using still images (PC-Stills)
- A HMD version using still images (HMD-Stills)
- A HMD version using videos (HMD-Videos)

The inclusion of both the HMD-Stills and HMD-Videos conditions will enable us to understand whether any changes in performance are the result of an increased field of view, or whether the more immersive and ecologically valid environment provided by videos can improve trainees understanding even further. Each of these adaptations will be described in more detail in the Method section below.

2. Method

2.1 Participants

76 participants were recruited, with three dropping out for practical reasons. Therefore, the results from 73 participants (28 male, 45 female) will be reported. Participant age ranged from 18.23 to 50.21 years ($M=27.18$, $SD=7.69$), with a one-way analysis of variance (ANOVA) showing no significant age differences between the conditions ($F(3,67)=0.81$, $p=0.49$). There were a similar number of males and females in each group (see Table 1). All participants were non-drivers, who had not yet started the driver training process. This group were selected to ensure that there were no practice effects arising from previous experience of the HPT, as the materials were very similar to the test included in the UK licensing process

Table 1: Breakdown of Age & Gender across experimental groups.

Condition	Gender		Age	
	Male	Female	Mean	SD
PC	7	11	28.83	7.53
HMD-Stills	7	12	27.79	9.67
HMD-Video	8	11	27.16	6.58
Control	6	11	24.86	6.49

Participants made one visit to the testing laboratory, lasting approximately 90 minutes, and they were paid £20 for their participation.

2.2 Training Scenarios

RAPT-UK consists of six scenarios, which were adapted from the schematics developed for RAPT 1 (Human Performance Lab, 2005), based on input from a workshop of road safety experts. The scenarios focused on latent risks associated with disrupted visibility either due to road geometry or the presence of another vehicle. Two of the scenarios focused on the challenges associated with overtaking a stopped **bus**, helping drivers to understand the potential for hidden oncoming traffic in the overtaking lane when moving to overtake a stopped bus (Bus1), and the potential risks associated with pedestrians disembarking the bus and attempting to cross the road when moving to re-enter the driving lane after overtaking a stopped bus (Bus2; see Figure 2). A **motorway** scenario focused on the need to constantly monitor merging lanes to be ready to accommodate vehicles entering the motorway. A sharp **bend** to the left, with limited visibility, was designed to encourage drivers to monitor the area to the far left to predict the approach of oncoming vehicles as early as possible. A **roundabout** scenario contained two potentially risky zones, a blind drive, and the roundabout itself. Drivers in this scenario would need to be aware of the potential for vehicles to emerge suddenly from areas where visibility was obstructed by either trees or buildings, while also being aware of the need to slow down and monitor traffic coming from the right on their approach to the roundabout. The final scenario was a junction with a **pedestrian crossing**, which focused on facilitating an understanding of the need to monitor for potentially hidden pedestrians approaching from either direction.

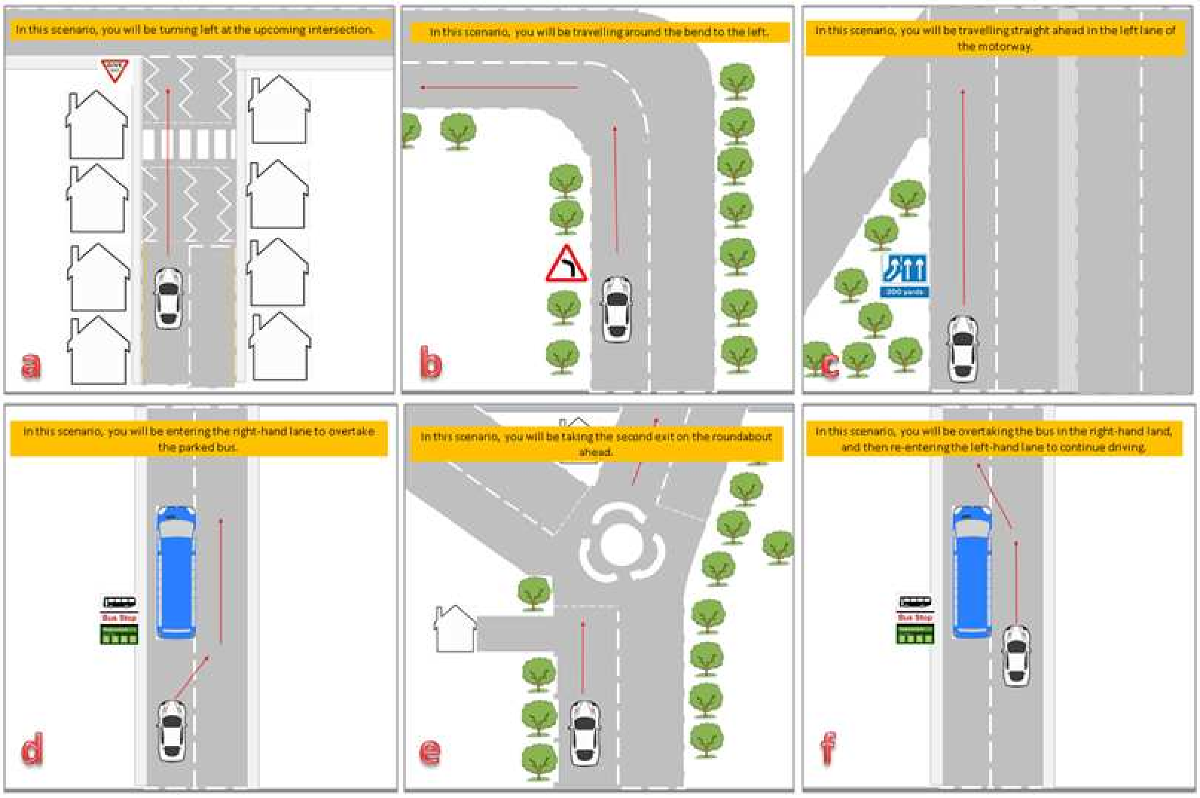

Figure 2: Schematic representations of the six RAPT scenarios: (a) pedestrian crossing, (b) bend, (c) motorway, (d) bus1, (e) roundabout, and (f) bus2

A series of locations in Leeds, UK were identified to capture the requirements of each selected scenario. The most appropriate clips for each location were chosen from a series of videos taken using a Nikon Keymission 360 camera, leading to a total of six clips, lasting in duration from 7 to 15 seconds. These clips were post-processed to remove any camera seaming, and to ensure that any identifying information such as faces and number plates were removed. Due to problems with playing 4K videos in MP4 format on Windows, the videos were converted to AVI files with highest quality settings. Playback of 2D and 3D videos was implemented using the Unity 3d gaming engine.

2.3 Apparatus

For the two conditions involving a Head Mounted Display (HMD), a HTC Vive HMD was used, which enables accurate head-tracking, and has a resolution of 1080 x 1200 per eye and a 110° field of view. The 360° videos and still images were displayed on the inside of a high-resolution sphere in a virtual 3D environment. The camera was positioned in the centre of the sphere, and could rotate, but not move forwards or backwards. This system incorporated head-tracking to allow scanning of the 360° frames, and participants were seated facing the active base station, using a hand-held clicking device to indicate when they believed they were looking at a risky area (see Figure 3).

Figure 3: HTC Vive Head Mounted Display and Clicker

For the PC-Stills condition, the 2D videos were shown on a Dell PC, and the aspect ratio of the videos matched the aspect ratio of the monitor. These 2D videos were played full screen, and individual frames from the video were used to show still images. The participants were instructed to click on the appropriate point in the scene where they would look, and could pan left and right before they chose a location.

For the PC-Stills and HMD-Stills conditions, a series of still images were extracted from the videos, with a 1.6 second gap between each of the images selected. Each frame was shown for approximately 3 seconds. Thus, the number of still images per scenario ranged from 5 to 8, with playback slowed by close to a factor of two. For the HMD-Video version, videos were played back to the participants at 0.66 of the recording rate, to enable users to have sufficient time to understand the road environment and identify areas of interest.

For both the PC and HMD solutions, an empty circle was displayed in front of the Unity camera that defined the direction of regard in the virtual world. The circle appeared in the centre of the screen on the PC, and directly in front of the participants head in the headset. While using the PC, participants clicked on the left mouse button to indicate that a potentially risky area was contained in the centre of the circle, while participants used the click controller to identify potentially risky areas in the HMD version. The circle was white when the mentioned buttons were not pressed, and blue when they were. The diameter of the circle was 4 degrees in the field of view of the camera, and was positioned 1 metre in front of it, ensuring that while wearing the headset the participants were aware of the circle and its position, but it did not draw their attention from the content displayed on the interior of the sphere, and also did not cover any part of the content.

2.4 Development of Training Content

The most important risk areas in each scenario were highlighted using the still images extracted from the original video materials (see Figure 4) and an algorithm was developed to identify when participants clicked within the relevant areas of interest. These areas of interest were reviewed by two subject matter experts at the University and by a member of the UK Department for Transport. Areas circled in red and blue denote the areas which the participant had to click. Participants could not see these circles during training. For the HMD-video condition, the same areas of interest were used. The areas identified in the still images were marked on the video for a period of 0.8 s before the still image appeared and 0.8 s afterwards to provide a continuous measurement of clicks in the risky areas.

Figure 4: Example of still image from RAPT-UK with risk areas circled in red & blue

The final step in the preparation of the training materials was the development of the training schematics, which provided visual guidance and text-based explanations about the inherent risks within a scenario, the location of these risks, and how to manoeuvre the scenario. The development of these training schematics was closely based on the instructions provided for RAPT-3 (Pradhan et al., 2009). Figure 5 provides an example of a training schematic for the pedestrian scenario, which corresponds to the risky areas identified in Figure 4. For the HMD groups, all of the schematics, images, and videos were viewed through the HMD, while the PC group viewed the schematics on the computer monitor. The HMD materials consisted of virtual 2D images, which were very similar to the images displayed on the PC monitor.

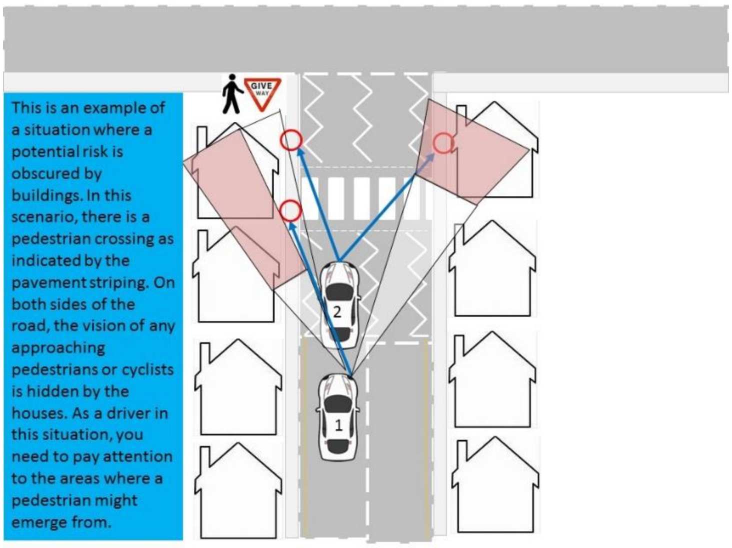

Figure 5: Example of training schematic for the Pedestrian scenario

The RAPT-UK pre- and post-training tests were used to evaluate participants' performance improvements after the training. In addition, all participants performed two versions of the Driving Vehicles Standards Association HPT – before and after RAPT-UK training. Each of the HPT versions consisted of 10 video clips, with 11 hazards in each test. A score of 58.66% was required to pass the test. The order of the HPTs was fully counterbalanced across participants, with no videos repeated across versions.

2.5 Experimental Design

On arrival to the experimental lab, participants were assigned to one of four training groups:

- PC-Stills
- HMD-Stills
- HMD-Videos
- Control Group

A between-subjects design was used, with the overall performance of the four experimental groups being compared. An α-value of .05 was used as the criterion for statistical significance, and partial eta-squared was computed as the effect size statistic. As the trainees in this experiment were all non-drivers of different ages, and research has shown that age and experience both make unique contributions to crash risk on the road (Forsyth et al., 1995), where possible, age was included as a covariate in the analysis. Degrees of freedom were Greenhouse-Geiser corrected when Mauchly's test showed a violation of sphericity.

2.6 Procedure

Upon arrival, participants were briefed on the experimental protocol. They were then shown an introductory video about the HPT (https://www.youtube.com/watch?v=SdQRkmdhwJs), prior to completing the first version of this test. Participants who were assigned to either the HMD-Stills or HMD-Video groups were then shown the HMD, and given an opportunity to practice looking around and clicking while wearing this headset. Participants could remove the headset for adjustment if they wished during the practice session, but once the training began, the HMD remained on until the post-training RAPT test was completed (approx. 30 minutes). Once they were comfortable with the headset, the RAPT file was loaded and they were asked to follow the instructions appearing on the screen. The RAPT training consisted of three main sections: (1) a Pre-Test, (2) Training, and (3) a Post-test. In the pre- and post-test sections, participants were presented with the still images/videos for each driving scenario, and were required to click on the areas where they would look if they were driving the scenario. During the training, participants were shown the top-down schematics of each of the scenarios and were then presented with the still images/videos of the scenario. If participants did not click on the correct areas of the scene, they were asked to re-review the schematics, and then to look at the images/video of the scene once more. If a participant made six failed attempts at a given scenario, the experimenter made the risky areas visible during the next playback. Participants in the control group completed an activity battery lasting approximately 30 minutes. This included two rounds of both a visual task, the arrows task (Jamson & Merat, 2005), and a non-visual 1-back task (Mehler et al., 2011). Once the training/activity battery was completed, all participants then completed a second version of the HPT.

3. Results

Due to an error in the application of the pass-rate algorithm for the bend scenario, it was not possible to compare the video version of RAPT to the still image versions. Therefore, the bend scenario has been excluded from the overall analysis of RAPT training and test performance.

3.1 Training Experience

In order to understand the effects of training modality on the difficulty of the training experience, a 3-way mixed between-within groups ANCOVA was conducted with Attempts (number of presentations required before all relevant risks were identified in a given scenario) as the dependent variable, and Scenario (Bus 1/Motorway/Pedestrian Crossing/Bus 2/Roundabout) and Condition (PC-Stills/HMD-Stills/HMD-Videos) as the independent variables, with age as a covariate. As the participants in the control group did not complete any of the RAPT-UK materials, they were not included in this analysis.

There was no significant main effect of age ($F(1,50)=1.18$, $p=0.28$). There was a significant effect of Condition ($F(2, 50)=39.71$, $p<0.001$, $np^2=0.61$), with participants requiring significantly more attempts in the HMD-Video condition (M=3.86, SE=0.22) than in either the PC-Stills (M=1.49, SE=0.22, $p<0.001$) or HMD-Stills (M=0.48, SE=0.21, $p<0.001$) conditions. There was no significant effect of Scenario ($F(2.75, 137.31)=1.17$, $p=0.33$), but there was a significant interaction between Scenario and Condition ($F(5.49, 137.31)=8.39$, $p<0.001$, $np^2=0.25$).

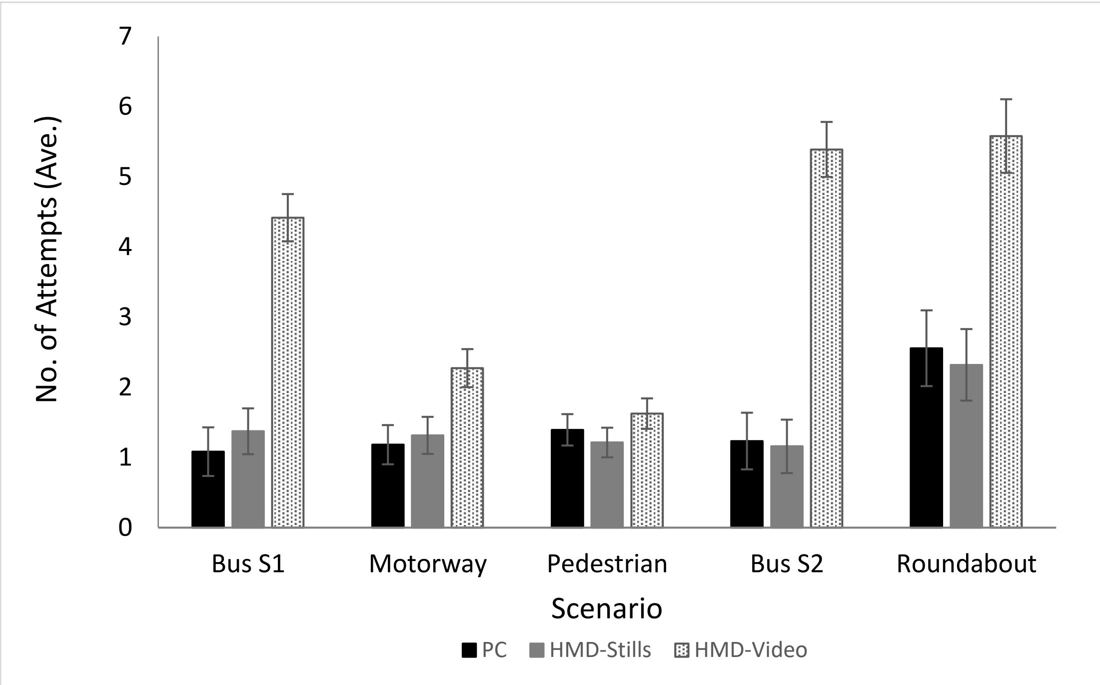

Figure 6: Interaction effect of Experimental Condition and Scenario Type on number of training attempts required

As Figure 6 shows, participants in the HMD-Video condition needed more attempts to understand the risky locations in the bus and roundabout scenarios than participants in the other groups. A series of one-way ANOVA's comparing each of the individual scenarios for the three experimental conditions found a significant difference for the Bus1 ($F(2,53)=27.26$, $p<0.001$), Motorway ($F(2,53)=4.67$, $p<0.05$), Bus2 ($F(2, 53)=38.53$, $p<0.001$), and Roundabout scenarios ($F(2,53)=11.05$, p<0.001), but not for the Pedestrian scenario ($F(2,53)=0.78$, $p=0.47$). Post-hoc tests showed that in each case, participants in the HMD-Video condition took significantly more attempts than either of the other conditions.

A series of repeated-measures ANOVAs were also conducted to evaluate the effect of Scenario on the number of Attempts for each condition. Results for the PC-Stills ($F(1.47, 25.03)=7.31$, $p<0.001$, $np^2=0.30$) and HMD-Stills groups ($F(1.56, 28.06)=5.72$, $p<0.05$, $np^2=0.24$) indicated a significant effect of Scenario, with participants taking significantly more attempts to pass the Roundabout scenario than any of the others. For the HMD-video group ($F(2.58, 46.35)=14.822$, $p<0.001$, np2=0.45), the number of attempts varied significantly between scenarios, with participants taking significantly more attempts to pass the Roundabout and Bus scenarios than any of the others.

3.2 Improvement in RAPT-UK Scores

In order to evaluate whether or not the different implementations of RAPT-UK were successful in improving learners' ability to understand the situations they were trained on, a 3-way mixed between-within groups ANCOVA was conducted, with RAPT-UK Score (percentage of correctly identified risk areas) as the dependent variable, and Test-Timing (pre-/post-training), Scenario (Bus 1/Motorway/Pedestrian Crossing/Bus 2/Roundabout) and Condition (PC-Stills/HMD-Stills/HMD-Video) as the independent variables, with age as a covariate.

There was a significant effect of age ($F(1,50)=15.75$, $p<0.001$, $\eta p^2= 0.24$),which was controlled for in all other analyses. To further explore this result a series of Pearson's correlation coefficients were calculated to investigate the relationship between age and overall RAPT scores in the pre- and post-training tests. Results indicated that there were no significant relationships between age and RAPT performance for the pre-training test ($r = -0.23$, $p = 0.09$). However, age was significantly negatively correlated with each of the scenario scores for the post-training test ($r = -0.51$, $p<0.001$). This suggests that the effectiveness of the training may have decreased as age increased.

There was a significant main effect of Condition ($F(2,50)=4.08$, $p<0.05$, $\eta p^2= 0.14$), with participants in the PC-Stills condition (M=42.68, SE=3.06) identifying significantly more hazardous areas across the two tests than participants in the HMD-Video condition (M=30.49, SE=2.97, $p<0.05$). There was a significant improvement in scores between the pre- (M=19.88, SE=1.78) and post-training tests (M=53.52, SE=2.14; $F(1,50)=66.07$, $p<0.001$, $\eta p^2= 0.57$). However, there was no significant interaction between Test-Timing and Condition ($F(2,50) = 1.01$, $p=0.37$). Figure 7 shows that scores were lower in the HMD-video version than in the other two versions of RAPT-UK. Nonetheless, it is clear that all three training regimes were successful in improving learners' ability to identify the risky areas within the RAPT-UK scenarios.

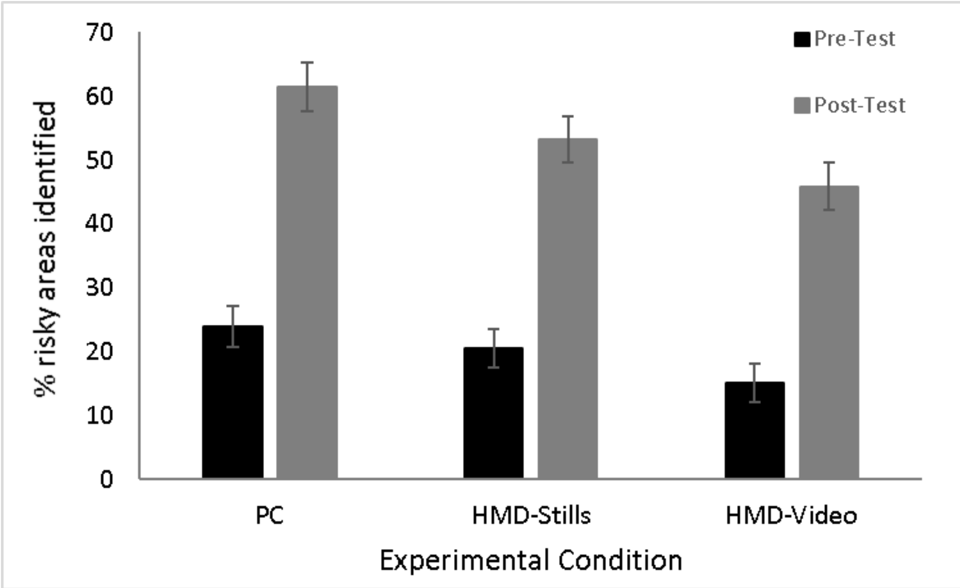

Figure 7: RAPT-UK pre- and post-training scores for the 3 experimental conditions

There was no significant effect of Scenario (F(4, 200)=0.39, p=0.82), but there was a significant interaction between Condition and Scenario (F(8,200)=7.45, p<0.001, ηp^2=0.23). As Figure 8 shows, participants in the PC-Stills group were significantly more successful than participants in either the HMD-Stills (p<0.05) or HMD-Video (p<0.001) groups at identifying risky areas in the first bus scenario (F(2,53)=10.42, p<0.001). Participants in the HMD-Video group identified significantly fewer hazards than the PC-Stills (p<0.01) or HMD-stills (p<0.05) groups in the second bus scenario (F(2,53)=7.45, p<0.01). However, the HMD-video group identified significantly more hazards than the PC-Stills group in the motorway scenario (F(2,53)=3.86, p<0.05). There were no significant group differences for either the pedestrian (F(2,53)=1.33, p=0.27) or roundabout (F(2,53)=1.92, p=0.16) scenarios. A series of repeated-measures ANOVAs show that the performance of the HMD-stills group was relatively consistent across scenarios (F(4,72)=1.90, p=0.12), while the performance of the HMD-video group varied significantly depending on the scenario (F(4,72)=12.49, p<0.001, np^2=0.41). This group was significantly weaker at identifying risky areas in the second bus scenario, while being significantly stronger at identifying risky areas in the motorway scenario. The performance of the PC-Stills group also varied significantly across the scenarios (F(4,68)=7.92, p<0.001, np^2=0.32), with participants showing significantly better understanding of the first bus scenario than any of the other scenarios.

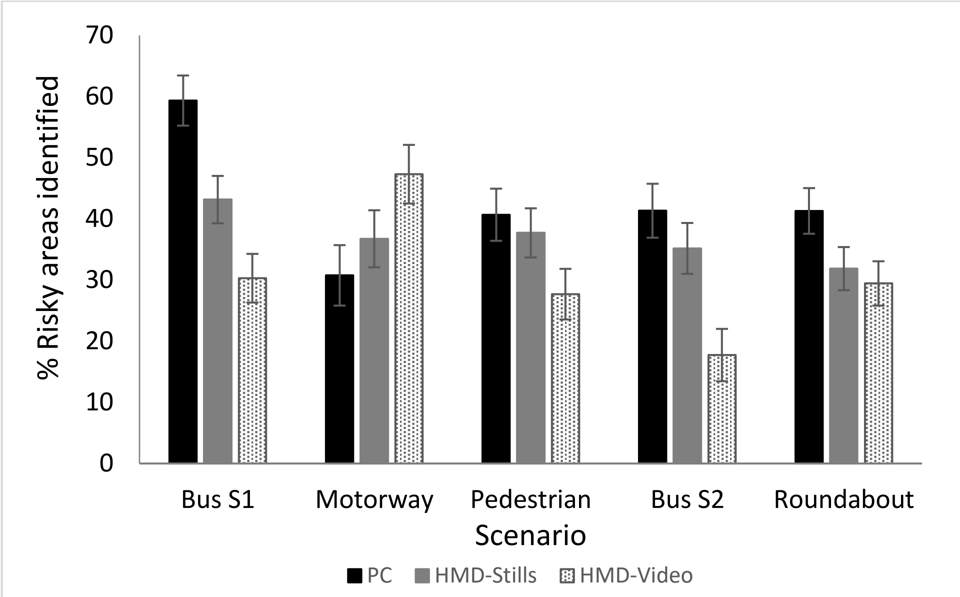

Figure 8: RAPT-UK interaction between scenario type and experimental condition

3.3 Hazard Perception Performance

In order to evaluate whether or not the different implementations of RAPT improved learners' performance in a traditional measure of hazard perception, a 2-way mixed between-within groups ANCOVA was conducted with HPT Score as the dependent variable, and Test-Timing (pre/post-training) and Condition (PC-Stills/HMD-Stills/HMD-Videos/Control) as the independent variables, and age as a covariate.

There was a significant effect of age ($F(1,66)=6.89$, $p<0.01$, $np^2=0.09$), which was controlled for in all other analyses. To explore this effect further, Pearson's correlations were calculated. The results show that there was no significant relationship between age and pre-training HPT scores ($r = -0.193$, $p=0.107$). However, there was a significant relationship between age and post-training HPT scores ($r=-0.292$, $p=0.014$), suggesting that HPT scores decreased with age. This implies that the training may have been more successful for younger participants.

There was no significant effect of Test-Timing ($F(1, 66)=3.01$, $p=0.09$), or Condition ($F(3, 66)=1.12$, $p=0.35$). However, there was a significant interaction effect ($F(3, 66)=5.10$, $p<0.01$, $\eta p^2=0.19$; see Figure 9). A series of post-hoc paired samples t-tests showed no significant change in performance for the PC-Stills ($t(17)=0.62$, $p=0.55$), HMD-Stills ($t(18)=-1.83$, $p=0.08$), or Control ($t(16)=1.65$, $p=0.12$) groups, but there was a significant improvement in performance for the HMD-Video group ($t(18)=-3.04$, $p<0.01$) from the pre- to post-training HPT. A series of one-way ANOVAs found no significant differences between the Conditions in either their pre-training ($F(3,69)=0.51$, $p=0.68$) or post-training results($F(3,69)=1.48$, $p=0.23$), suggesting that any changes in performance resulted from the experimental manipulation and not between-group differences.

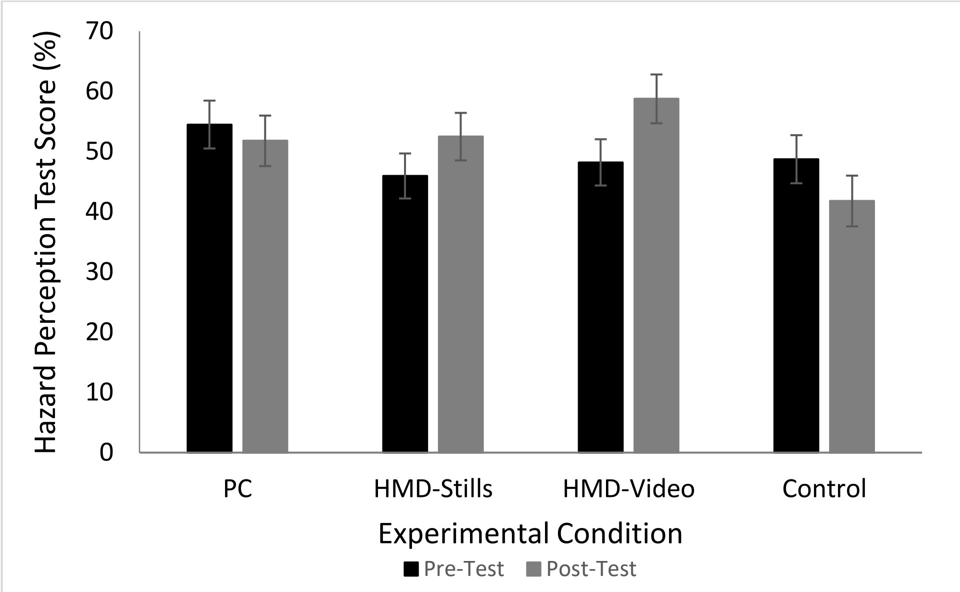

Figure 9: Hazard Perception Test scores across experimental groups

As the RAPT focuses on training drivers to scan the environment for potentially risky scenarios (environmental hazards) rather than developing hazards (behavioural hazards), it was anticipated that the training would lead to a greater number of clicks outside of the specified hazard windows. This was evaluated by examining the total number of clicks made per scenario in the two HPTs. It should be noted that a cheating detection algorithm was implemented to capture any continuous clicking behaviours, providing confidence that any clicks were linked to perceived risks.

A 2-way mixed between-within groups ANCOVA was conducted with Total Clicks as the dependent variable, and Test-Timing (pre-/post-training) and Condition (PC-Stills/HMD-Stills/HMD-Videos/Control) as the independent variables, and age as a covariate. Results indicate that there was no significant effect of age (F(1,66)=1.52, p=0.22). There was a significant effect of Condition (F(3, 66)=2.95, p<0.05, np^2=0.12), with participants in the Control condition (M=34.73, SE=5.30) making significantly fewer clicks than participants in the PC-Stills condition (M=55.90, SE=5.27, p<0.05) across the two tests. There was no significant effect of Test-Timing (F(1, 66)=0.89, p=0.35), but there was a significant interaction between Test-Timing and Condition (F(3, 66)=4.60, p<0.01, ηp^2=0.17). As Figure 10 shows, the Total Clicks increased after training in all three RAPT groups, but decreased for the control group. A series of post-hoc paired samples t-tests showed that there was no significant change in performance for the PC-Stills (t(17)=-1.67, p=0.12), or Control group (t(16)=1.54, p=0.14) conditions, but there was a significant increase in clicks for the HMD-Stills (t(18)=-2.66, p<0.05) and HMD-Video groups (t(18)=-3.95, p<0.01) from the pre- to post-training HPT. A series of one-way ANOVAs found no significant differences between the Experimental Conditions in pre-training (F(3,69)=1.66, p=0.18), or post-training results(F(3,69)=2.52, p=0.07), suggesting that any changes in performance were due to the experimental manipulation and not between-group differences.

Figure 11 provides an example of the spread of clicks within one section of one video from the HPT. It compares the number of clicks made by participants who had completed this section prior to receiving any training, to the number of clicks made by participants who completed it after training. It is very clear by the increase in blue, yellow, and orange stars that all three RAPT training regimes led to participants more regularly identifying areas of the environment which might contain potential risks, while the control group showed no such change in behaviour.

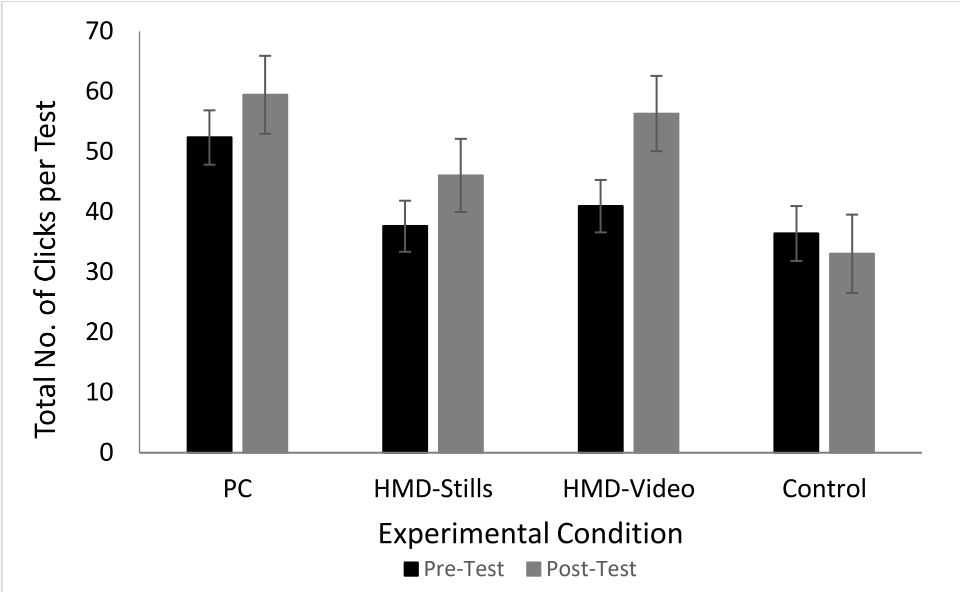

Figure 10: Total number of participant clicks in pre- and post-training HPT

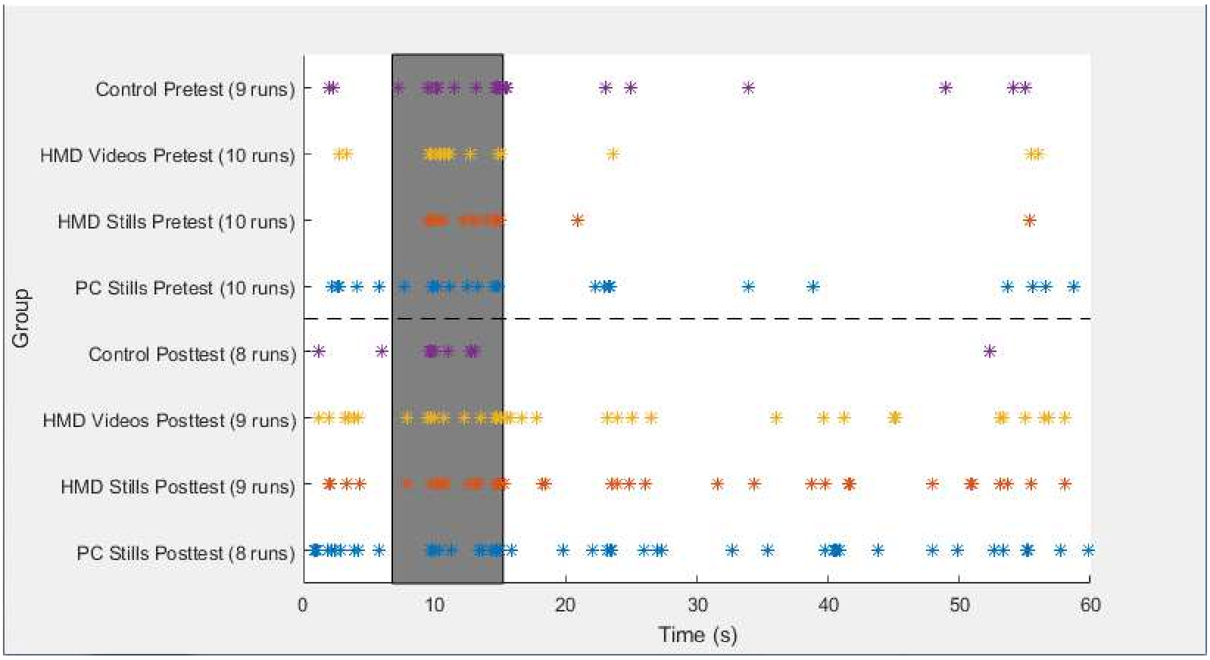

Figure 11: Spread of participant clicks across a sample hazard window in the pre- and post-training HPT (the order of the tests was counterbalanced so this is a between groups comparison)

4. Discussion

The first aim of this study was to evaluate an adaptation of RAPT3 for the UK context. Overall, the results indicate that this adaptation was successful, with performance improvements arising in both the post-training RAPT-UK test and HPT. To gain a deeper understanding of the impact of modality on the training experience, the number of attempts required to pass a training scenario in both PC and HMD versions of RAPT-UK were compared. Results suggest that participants in the HMD-video condition had the most difficulty in completing the training, requiring more attempts to pass scenarios than either of the other training groups. However, despite differences in the length of time taken to complete the training, all three training regimes led to an improvement in the ability to identify potentially hazardous areas of the RAPT-UK driving scenarios. Indeed, it is possible that the increased difficulty experienced by participants in the HMD-video condition during training may have led to a deeper processing of materials and a greater ability to transfer learning, as evidenced by the improved performance of this group in the HPT. Agrawal et al. (2018) found that participants in their HMD-video group (V-RAPT) anticipated a greater proportion of environmental hazards than the RAPT-PC group. The results of the current study suggest that this discrepancy was not due to the use of videos per se, or a result of the increased validity provided by more realistic head movements. Instead, it is possible that this result was linked to the differences in training content between V-RAPT and RAPT-PC. Since several studies have shown that there is a potential for simulator sickness while using HMDs (e.g. Benz, Riedl, & Chuang), this is an important finding, as it suggests there is no difference between the use of still images and videos, at least in terms of performance on the RAPT tests.

Previous research has shown that different hazards may be more/less successful in discriminating between experience groups (Sagberg & Bjornskau, 2006; Crundall et al., 2012). The analysis of the specific training scenarios within RAPT-UK provides support for this idea, by showing that participants' understanding of where to look was influenced by the characteristics of the environment. Training modality effects varied depending on the type of scenario being explored, with participants in the PC-Stills group being particularly good at identifying where to look when approaching a parked bus, while participants in the HMD-Video group showed the most effective scanning behaviour in the motorway condition. The fact that differences emerged between all three modalities suggests that both field of view (PC vs HMD) and presentation modality (still images vs video) have an impact on how RAPT training is processed. This suggests that different versions of RAPT may be more or less suitable, depending on the desired training materials. For example the HMD-video version may be more suitable for more high-speed situations such as motorway driving, where the PC or HMD-stills versions may be more appropriate in complex slow-moving situations such as overtaking a parked bus. It is also possible that different implementations of RAPT-UK may be more appropriate at different stages of the learning process. Skill acquisition research suggests that expert performance occurs when a learner moves from declarative (conscious, verbal, knowledge) to procedural knowledge (automatic performance) of a particular skill (Anderson , 1992). The results of the current study suggest the possibility of using different modalities to enhance the skill acquisition and knowledge compilation process by, for example, using the PC-based version as a starting point for novice drivers, who could then move onto the more difficult HMD-video version once they have acquired the basic declarative knowledge required to understand hazards in a static environment.

An analysis of participants' performance on the UK Department for Transport HPT showed that participants in the HMD-Video condition significantly improved after training, while the HMD-Stills, PC-Stills and Control groups did not. This suggests that the mode of delivery had an impact on trainees' ability to detect emerging hazards, with participants presented with videos in both the training and testing contexts, showing an improvement in performance after training, compared to those presented with still images. As RAPT focuses on training drivers to scan the environment for potentially risky scenarios (environmental hazards) rather than reacting to developing hazards (behavioural hazards), it was anticipated that the training would lead to participants identifying additional potentially hazardous areas where the hazard didn't materialise. An analysis of clicks made showed that participants in the three training conditions identified more potential hazards in the HPT materials after training. A comparison of the timings of the response patterns shows that these additional clicks were clustered in similar places (see Figure 11), suggesting that they were a result of understanding potentially hazardous zones, rather than random clicking throughout. This suggests that the scanning patterns learned during training transferred successfully to scenarios not encountered as part of the training process, and may suggest that the HPT could be adapted to also provide a measure of environmental hazard anticipation.

Taken together, these results show that the RAPT-UK can be considered as an engaging, and effective training tool for improving novice drivers' scanning patterns. Although previous implementations of RAPT had focused on young drivers (e.g. Thomas et al., 2016), the current study adopted a wider age range. However, the results suggest that the effectiveness of the training decreased with age, implying that this tool may be most suitable for young, novice drivers. The incorporation of a wider field of view through the HMD-still and HMD-video versions was also successful, with the greatest improvements arising for the HMD-video training. This study provides a successful proof of concept with a small sample size. It should be noted that all participants were non-drivers with no driving experience, which may have impacted the results. However, the fact that the training was successful in improving this group's performance provides reassurance that it would be beneficial for participants who already have some understanding of driving hazards, and it is likely that the effects would be stronger for trainee drivers. Further research is needed to gain an understanding of other scenarios which could be incorporated to these training materials. Another consideration is that the current study required participants to actively click on risky areas in both the PC and HMD conditions. It would be interesting to explore whether the use of an integrated eye-tracker enabling an evaluation of fixations would enhance the training effectiveness (see Bozkir et al., 2019; Pai Mangalore et al., 2019), or whether the conscious act of clicking on a risk area is required for participants to rapidly develop risk awareness skills. This type of evaluation would also enable an understanding of whether the driver had consciously decided not to click on some areas they had looked at. Finally, future research should explore the use of driving simulator and on-road evaluations to provide a more in-depth understanding of the long-term implications of training modality in terms of increasing road safety in the UK.

Acknowledgements

This project was funded by the UK Department for Transport through the Innovation Challenge Fund. The authors would also like to thank Dr. Anuj Pradhan for his insights into the development of the original RAPT training programme, and Hrvoje Jukic for his programming assistance.

References

Agrawal, R., Knodler, M., Fisher, D.L., & Samuel, S. (2018). Virtual reality headset training: Can it be used to improve young drivers' latent hazard anticipation and mitigation skills. *Transportation Research Record*, 2672, 20-30.

Anderson, J. R. (1992). Automaticity and the ACT theory. *The American Journal of Psychology*, 105(2), 165-180.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612-637.

Benz, T.M., Riedl, B., & Chuang, L.L. (2019). Projection displays induce less simulator sickness than Head-Mounted Displays in a real vehicle driving simulator. *In 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '19)*, September 21–25, 2019, Utrecht, Netherlands. ACM, New York, NY, USA. https://doi.org/10.1145/3342197.3344515

Bozkir, E., Geisler, D., & Kasneci, E. (2019, September). Assessment of Driver Attention during a Safety Critical Situation in VR to Generate VR-based Training. In ACM Symposium on Applied Perception 2019 (p. 23). ACM.

Crundall, D., Chapman, P., Trawley, S., Collins, L., van Loon, E., Andrews, B., & Underwood, G. (2012). Some hazards are more attractive than others: Drivers of varying experience respond differently to different types of hazard. *Accident Analysis & Prevention*, 45, 600-609.

Fisher, D. L., Pradhan, A.K., Hammel, K.R., DeRamus, R., Noyce, D.A., & Pollatsek, A.P. (2003). Are younger drivers less able than older drivers to recognise risks on the road? *Injury Insights*, February/March, 1, 2, 7.

Forsyth, E., Maycock, G., & Sexton, B. (1995). Cohort study of learner and novice drivers: Part 3, accidents, offences and driving experience in the first three years of driving. *TRL Project Report 111*. Crowthorne, UK: TRL.

Grayson & Sexton,(2002).The development of hazard perception testing. *TRL Report 558*. Crawthorne, UK: TRL.

Groeger, J.A., & Banks, A.P. (2007). Anticipating the content and circumstances of skill transfer: Unrealistic expectations of driver training and graduated licensing? *Ergonomic*s, 50, 1250-1263.

Horswill, M. S., & McKenna, F. P. (2004). Drivers' hazard perception ability: Situation awareness on the road. In S. Banbury & S. Tremblay (Eds.), *A Cognitive Approach to Situation Awareness: Theory and Application (pp. 155-175)*. Hampshire, England: Ashgate.

Human Performance Lab (2005). *Risk Awareness and Perception Training (RAPT) - Version 1: Training Program for Younger Drivers*. University of Massachusetts: Amherst. Accessed at http://www.ecs.umass.edu/hpl/RAPT/RAPT1_Full_Version_Jun08.pdf (30th May 2019)

Jackson, L., Chapman, P., & Crundall, D. (2009). What happens next? Predicting other road users' behaviour as a function of driving experience and processing time. *Ergonomics*, 52(2), 154-164.

Jamson, A.H., & Merat, N. (2005). Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated driving. *Transportation Research Part F*, 8, 79-96.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis, *Journal of Applied Psychology*, 93, 59-69.

Pai Mangalore, G., Ebadi, Y., Samuel, S., Knodler, M. A., & Fisher, D. L. (2019). The Promise of Virtual Reality Headsets: Can They be Used to Measure Accurately Drivers' Hazard Anticipation Performance? *Transportation Research Record*, 0361198119847612.

McCartt, A. T., Shabanova, V. I., & Leaf, W. A. (2003). Driving experience, crashes and traffic citations of teenage beginning drivers. *Accident Analysis & Prevention*, 35(3), 311-320.

McDonald, C.C., Goodwin, A.H., Pradhan, A.K., Rosomer, M.R.E., & Williams, A.F. (2015). A review of hazard anticipation training programs for young drivers. *Journal of Adolescent Health*, 57, 15-23.

McKenna, F. P., Horswill, M. S., & Alexander, J. L. (2006). Does anticipation training affect drivers' risk taking? *Journal of Experimental Psychology: Applied*, 12(1), 1-10.

Mehler, B., Reimer, B., & Dusek, J. (2011). MIT AgeLab delayed digit recall cask (n-back). Retrieved from http://agelab.mit.edu/system/files/Mehler_et_al_n-back-whitepaper_2011_B.pdf

Pollatsek, A., Narayanaan, V., Pradhan, A., & Fisher, D. (2006). Using eye movements to evaluate a PC-based risk awareness and perception training program on a driving simulator. *Human Factors*, 48, 3, 447 – 464.

Pradhan, A. K., D. L. Fisher, A. Pollatsek, M. Knodler, and M.Langone. Field Evaluation of a Risk Awareness and Perception Training Program for Younger Drivers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50, No. 22, 2006, pp. 2388–2391.

Pradhan, A. K., Pollatsek, A., Knodler, M., & Fisher, D. L. (2009). Can younger drivers be trained to scan for information that will reduce their risk in roadway traffic scenarios that are hard to identify as hazardous? *Ergonomics*, 52(6), 657-673.

Sagberg, F., & Bjørnskau, T. (2006). Hazard perception and driving experience among novice drivers. Accident Analysis & Prevention, 38(2), 407-414.

Taylor, T., Masserang, K., Pradhan, A., Divekar, G., Samuel, S., Muttart, J…..Fisher, D. (2011). Long-term effects of hazard anticipation training. *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design 2011* (pp. 187-194). Iowa City: Public Policy Centre, University of Iowa.

Thomas, F.D., Rilea, S.L., Blomberg, R.D., Peck, R.C., & Korbelak, K.T. (2016). Evaluation of the safety benefits of the risk awareness and perception training program for novice teen drivers. *Report No. DOT HS 812235*. Washington DC: National Highway Traffic Safety Administration.

Unverricht, J., Samuel, S., & Yamani, Y. (2018). Latent hazard anticipation in young drivers: Review and meta-analysis of training studies, *Transportation Research Record*, 2672, 11-19.

Wells, P., Tong, S., Sexton, B., Grayson, G. and Jones, E. (2008) Cohort II: A Study of New Drivers. Volume 2 – Questionnaires and Data Tables. *Road Safety Research Report No. 81*. London: Department for Transport. Available online at: http://www.dft.gov.uk/pgr/roadsafety/research/rsrr/theme2/cohort.

Yamani, Y., Biçaksiz, P., Palmer, D.B., Cronauer, J.M., & Samuel, S. (2017). Following expert's eyes: Evaluation of the effectiveness of a gaze-based training intervention on young drivers' latent hazard anticipation skills. *Proceedings of the Ninth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, June 26-29, 2017, Manchester Village, Vermont.

Zhang, T., Li, J., Thai, H., Zafian, T., Samuel, S., & Fisher, D. (2016). Evaluation of the effect of a novice driver training program on crashes and citations. *Proceedings of the Human Factors and Ergonomics Society.* Washington DC: Human Factors and Ergonomics Society.